



University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Alex Noel Joseph Raj and Richard C. Staunton

Article Title: Rational filter design for depth from defocus

Year of publication: 2012

Link to published article:

<http://dx.doi.org/10.1016/j.patcog.2011.06.008>

Publisher statement: "NOTICE: this is the author's version of a work that was accepted for publication in Pattern Recognition. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Pattern Recognition, VOL:45, ISSUE:1, January 2012, DOI: 10.1016/j.patcog.2011.06.008

Rational Filter Design for Depth from Defocus

Alex Noel Joseph Raj¹ and Richard C. Staunton²

¹School of Engineering, University of Warwick, Coventry CV4 7AL, UK.

Email: A.N.Joseph-Raj@warwick.ac.uk

²School of Engineering, University of Warwick, Coventry CV4 7AL, UK.

Email: R.C.Staunton@warwick.ac.uk

Corresponding author: R.C. Staunton. Phone: +44 2476 523980, Fax: +44 2476 418922

This paper was published in the Journal: Pattern Recognition. The full reference is:-
Alex Noel Joseph Raj and Richard C. Staunton, (Jan. 2012) *Rational Filters Design for Depth from Defocus*. Pattern Recognition, Vol. 45 (No. 1). pp. 198-207.

Abstract

The paper describes a new, simple procedure to determine the rational filters that are used in the depth from defocus (DfD) procedure previously researched by Watanabe and Nayar [4]. Their DfD uses two differently defocused images and the filters accurately model the relative defocus in the images and provide a fast calculation of distance. This paper presents a simple method to determine the filter coefficients by separating the M/P ratio into a linear and a cubic error correction model. The method avoids the previous iterative minimisation technique and computes efficiently. The model has been verified by comparison with the theoretical M/P ratio. The proposed filters have been compared with the previous for frequency response, closeness of fit to M/P, rotational symmetry, and measurement accuracy. Experiments were performed for several defocus conditions. It was observed that the new filters were largely insensitive to object texture and modelled the blur more precisely than the previous. Experiments with real planar images demonstrated a maximum RMS depth error of 1.18% for the proposed, compared to 1.54% for the previous filters. Complicated objects were also accurately measured.

Key Words: Depth from defocus, M/P ratio, Rational filters, 3D imaging.

1. Introduction

This paper reports improvements to the accuracy of a well known Depth from Defocus (DfD) recovery algorithm. The algorithm is particularly suited to applications that require high-speed

processing. DfD is a technique in which the ‘defocus’ at a pixel in an image is used to estimate the distance from the lens to the corresponding point on an object. The method requires two differently focused images acquired from a single view point using a single camera, and the relative blur between the images is used to determine the in-focus axial points of each pixel and hence depth. In this way it differs from the allied method of depth from focus which may use several images [1,2]. Most approaches consider blurring as a linear shift invariant process either in the frequency or spatial domain. The defocused image is modelled as the convolution of the focused image with the point spread function (psf) of the lens [3-8]. The blur information is retrieved by the deconvolution process either in the frequency or spatial domain, and then related to the actual distance using the appropriate depth model. These methods offer an advantage in terms of computation and simplicity in the implementation of the algorithm. An accuracy of 1.2% has been reported from experiments using images of real scenes [4], however it can be difficult to compare the techniques as the accuracy depends on the range over which the measurements were made, with a narrow range close to the camera giving better results. Additionally some researchers use local smoothing operators to remove noise from the final depth map that improve their accuracy figures. Other methods (often statistical) consider the blurring as a shift variant process and retrieve a unique depth value not only along the optical axis but also along the x and y directions of the scene under investigation. These methods can be accurate (1%) and are efficient since they simultaneously retrieve depth and the radiance of the scene, but may not be suitable for practical, real-time, purposes since they are based on error minimisation techniques which require extensive computations [9-14]. Video-rate processing is a requirement for 3D TV, and fast processing extends the use of DfD for robotics and production line applications. Efficient DfD computation methods have been proposed [4,15,16], however in this paper, since we are concerned with video-rate depth estimation for every pixel in the image, and passive illumination, we have chosen an approach based on rational filters [4] as detailed more fully below.

Figure 1. Telecentric DFD system

are $2e$ apart with i_f lying in between. Thus on the object side a working range is defined by the lens law. Previous DfD methods that were based on the frequency domain approach [3-5] estimated the depth by considering the amplitude ratio of two defocused images at a particular radial frequency. Watanabe and Nayar [4] provided an improvement with the M/P ratio curve set that effectively models the lens defocus performance for particular working conditions at different texture frequencies. They considered the amplitude ratio between the difference (M) of the defocused images to their sum (P), and developed a set of broadband filters that modelled the curves (Figure 2a). Although the filters were designed in the frequency domain, the algorithm was implemented in the spatial domain by employing 2D convolutions. Only 7×7 filter kernels were required and the algorithm enabled the filters to be applied in parallel resulting in efficient hardware implementation. The main advantages of this method were: (1) Higher accuracy in depth estimation. The RMS error reported was 1.2% with respect to distance (which was better than any comparable methods [17]); (2) Invariance to scene texture and illumination (the depth detection error was less than 1% with respect to texture frequency [4]); and (3) The feasibility of high speed calculation using a hardware implementation. Additionally they researched telecentric optics [18] to enable the near and far-focused images to be optically registered by removing the magnification caused by refocusing the lens. In Figure (1) the telecentric aperture is labelled T and is at the front focal point of the lens. We have used telecentric optics with a Phase Correlation technique [19] to measure the radial shifts due to magnification, and also to optimally position the external aperture.

Despite the efficient algorithm, the main drawback of the previous method was the complicated design procedure based on an iterative minimisation technique to model the rational filters for any given defocus condition, and texture frequency. Watanabe [4] has published filter coefficients for the single defocus condition of 2.307 pixels, and subsequent researchers have only published work which used these and no other. In this paper we report a new, simple procedure for rational filter design, the Two Step Polynomial Approach [17]. We have compared our model with Watanabe and Nayar's, for the defocus condition of 2.307 pixels, for the accuracy with which the filters approximate the M/P ratio curves, and for the overall depth estimation accuracy. For the Two Step Polynomial Approach we have also provided the overall depth estimation accuracy for a range of defocus conditions, using both simulated and real images.

Section 2 describes how the Two Step Polynomial Approach was employed to model the rational filters for any given defocus condition. Step1 involved modelling the linear filters by fitting a linear model to the theoretical M/P ratio for each radial frequency, and Step2 determined a correction filter by computing the error between the theoretical and the linear models, and fitting a cubic function to it. Section 3 contains a comparison between the previous filters and the new ones. This has been done firstly with an analysis of the errors between the theoretical M/P curves and those modelled by the two sets of filters, and secondly by measuring the Fourier domain characteristics of each filter. Section 4 begins by outlining the depth estimation algorithm used. Then both simulated and real images have been used to estimate the accuracy of the depth estimation. The results show that the new filters estimate the depth map with a higher accuracy than the previous filters.

2. Filter design using the Two Step Polynomial Approach

This section describes the new procedure that uses a discrete M/P space and polynomials to determine the filter coefficients that were used in the DfD calculations. Three filters, Gm_l , Gp_l and Gp_2 were designed to model the lens defocus M/P ratio curves, and two pre-filters were used to remove dc (mean) values from the M and P input images. Since the model was based on the M/P ratios, these were calculated for a range of frequencies using the psf of the defocused lens. Various models of the psf exist where an impulse (delta) function can model a perfect lens in focus; a Gaussian function has been used for a lens in near focus [20]; and a pill box (cylindrical) function for a lens that is more defocused [3]. A single generalized Gaussian has been used to model all three regions [21], but to provide a meaningful comparison with Watanabe's filters [4] their pill box model has been used here. The psf was pre-computed using the Pillbox model for a range of normalized depth values, α from 0 to 0.99. So based on [4], the psf was modelled in the frequency domain using the equation

$$H(u, v) = H(u, v; z, F_e) = \frac{2F_e}{\pi z \sqrt{u^2 + v^2}} J_1\left(\frac{\pi z}{F_e} \sqrt{u^2 + v^2}\right) \quad \text{--- (1)}$$

Where, referring to Figure (1), $z = (1 \pm \alpha)e$ denotes the distance between the sensor plane and the focused image i_f . For a telecentric setup [17,18], $F_e = f/d$ represents the f -number of the lens, where d is the diameter of the aperture. $J_1(r)$ denotes the first order Bessel Function, and u, v are the frequencies in the horizontal and vertical directions respectively. In this case $H(u, v)$ is circularly symmetric so we define the radial frequency $f_r = \sqrt{u^2 + v^2}$. For the proposed filter design a discrete M/P space was required. To construct this for each lens, firstly α was discretized to take one of 11 equally spaced values within its positive range, and a psf was calculated for each of these. The continuous M/P ratio was as defined in [4]:

$$\frac{M}{P}(u, v; \alpha) = \frac{I_2(u, v) - I_1(u, v)}{I_2(u, v) + I_1(u, v)} \quad \text{---(2)}$$

Where $I_1(u, v) = I_f(u, v)H_1(u, v)$, $I_2(u, v) = I_f(u, v)H_2(u, v)$ and $I_f(u, v)$ is the Fourier transform of the in-focus image. I_1 and I_2 are therefore different images of the same scene that have been defocused by the transformed psfs H_1 and H_2 respectively. A set of 32 equally spaced texture frequencies, f_r , were considered in the range from minus to plus the folding frequency of 0.5 pixel^{-1} .

Substituting for I_1 , I_2 , and f_r we have $\frac{M(f_r; \alpha)}{P(f_r; \alpha)} = \frac{H(f_r; (1 - \alpha)e, F_e) - H(f_r; (1 + \alpha)e, F_e)}{H(f_r; (1 - \alpha)e, F_e) + H(f_r; (1 + \alpha)e, F_e)}$, which

is no longer a function of the image data. An 11×32 set of discrete M/P values were calculated for each of the values of α and f_r considered. Within this set the samples were equally spaced horizontally, but not vertically. An illustration of continuous M/P characteristics have been shown

in Figure 2(a), but the aim here was to construct a discrete set M/P samples as shown in Figure 2(b). The discrete M/P data set was central to the new filter design procedure presented below. To proceed with the filter design, the discrete M/P ratio was modelled as a linear combination of the three 2D filters, Gm_1 , Gp_1 and Gp_2 using a discrete version of the equation described by Watanabe [4].

$$\text{So, } \frac{M(f_r; \alpha)}{P(f_r; \alpha)} = \frac{Gp_1(f_r)}{Gm_1(f_r)} \beta + \frac{Gp_2(f_r)}{Gm_1(f_r)} \beta^3 \text{ ----} \quad (3)$$

This was then simplified to be a linear model plus an error correction model.

$$\frac{M(f_r; \alpha)}{P(f_r; \alpha)} = \frac{M'(f_r; \beta)}{P'(f_r; \beta)} + \frac{M''(f_r; \beta)}{P''(f_r; \beta)} \text{ ----} \quad (4).$$

Where $\frac{M(f_r; \alpha)}{P(f_r; \alpha)}$ represents the theoretical M/P ratio, $\frac{M'(f_r; \alpha)}{P'(f_r; \alpha)}$ the linear term: $\frac{Gp_1(f_r)}{Gm_1(f_r)} \beta$, and $\frac{M''(f_r; \alpha)}{P''(f_r; \alpha)}$ the error correction term: $\frac{Gp_2(f_r)}{Gm_1(f_r)} \beta^3$. Here α and β denote the actual and the estimated depth, and f_r is the radial texture frequency.

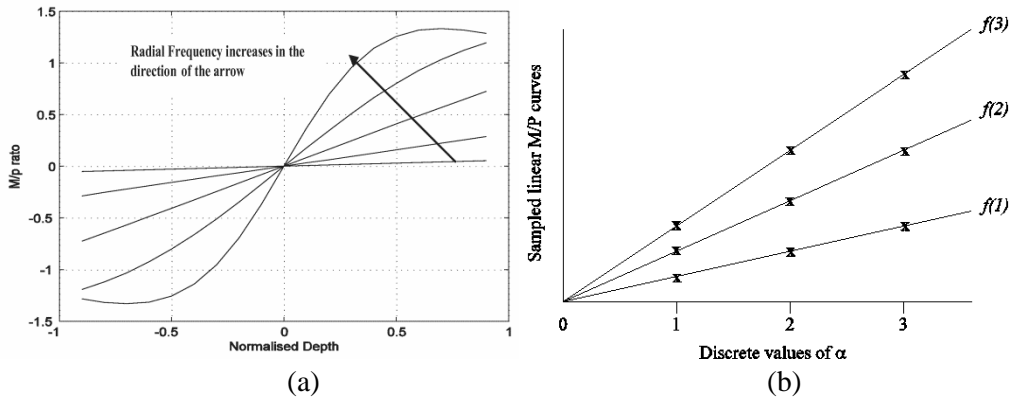


Figure 2. (a) Continuous M/P curves, (b) Discrete M/P ratio space

Now, consider the linear model: $\frac{Gp_1(f_r)}{Gm_1(f_r)} \beta$. It can be considered as a straight line passing through the origin, where all the curves converge, and the M/P ratio can be represented by the line: $y = Ax$, where the gradient for a particular radial frequency, $A(f_r)$, is equal to the ratio $\frac{Gp_1(f_r)}{Gm_1(f_r)}$.

Examining the example in Figure 2(a), the curves for the lowest two radial frequencies can be satisfactorily modelled by the linear model, but those for the higher frequencies diverge from this and so the cubic error term is included to enable a wider band of texture frequencies to be used in the depth calculation. The error term is described in Section 2.1. To proceed with the proposed design method a discrete set of M/P gradient functions were defined and used to calculate a prototype 1D FIR filter, G^P_{m1} . The Linear Model requires knowledge of:- (1) The gradient functions, $A(f_r)$, for each radial frequency; and (2) The given response of either G^P_{m1} or a prototype G^P_{p1} . To compute $A(f_r)$ the discrete ratio space was utilised since it provided a unique (within the allowable frequency range) M/P ratio for each normalized depth sample corresponding to a particular radial frequency. Hence to determine $A(f_r)$ at a radial frequency, f_{rj} , a linear function, $y=Ax$ was fitted to the M/P ratio for this frequency over the 11 discrete values of α .

Thus by considering the 32 radial frequencies, $A(f_r)$ was computed for each discrete frequency in the M/P space. $A(f_r)$ was then reduced to a 1D vector as the gradient for each frequency was constant for all values of α . $A(f_r)$ was then re-sampled to give 32 equally spaced samples. As $A(f_r)$ is the ratio of G^P_{p1} to G^P_{m1} it was possible to design prototype 1D versions of these filters. However, the frequency response of either G^P_{m1} or G^P_{p1} must be predefined. Since the required filter needs to possess a band-pass filter characteristic together with rotational symmetry [4], the response of G^P_{p1} was fixed as a Laplacian of Gaussian (LOG) [22] based on the equation

$$G^P_{p1} = \left(\frac{f_r}{f_{spread}} \right)^2 \exp \left(1 - \left(\frac{f_r}{f_{spread}} \right)^2 \right), \quad \text{---} \quad (5)$$

where $f_{spread} = 0.4 f_{nyquist}$. The constant 0.4 was used to ensure an acceptable width of the LOG filter. Note: to avoid any divide by zero problems, the second filter, G^P_{m1} was modelled as low-pass. Given G^P_{p1} and $A(f_r)$, the frequency response of the prototype G^P_{m1} was determined with ease as $G^P_{m1} = G^P_{p1}(f_r)/A(f_r)$. Now $Gm_1(f_r)$ is circularly symmetric, and so can be interpolated and reformulated from $G^P_{m1}(f_r)$ as a 32x32 response: $Gm_1(u,v)$. To obtain the 7x7 filter coefficients $gm_1(x,y)$, $Gm_1(u,v)$ was smoothed, re-sampled to 8x8, and then inverse Fourier transformed. The central 7x7 non-redundant coefficients were retained for the filter. The smoothing was performed by fitting a 12th order 2D polynomial to the 32x32 response. The same procedure was performed to obtain the coefficients of $gp_1(x,y)$ from the prototype $G^P_{p1}(f_r)$. Thus by employing just the linear model, the frequency responses of the filters Gp_1 and Gm_1 were found.

The pre-filters that prevent depth uncertainties due to dc propagation and depth ambiguity from high f_r M/P curves which peak within the range of α , needed to be band-pass, 2D, and rotationally symmetric. From [4] it was found that the LOG filter design of Gp_1 can also be used to design the pre-filters. To provide a smooth transition a spread factor of $f_{peak} = 0.74 f_{max}$ was used, where $f_{max} = 0.264 \text{ pixels}^{-1}$. The next section discusses the error correction model, and the frequency response of filter Gp_2 .

2.1. Error Correction Model

An additional filter: Gp_2 is required in the calculation of the cubic term in Equation (3). Gp_2 requires the same support as Gm_1 , Gp_1 and each pre-filter, and so increases the computational load by 25%. However as the cubic term allows the inclusion of high texture frequencies without compromising accuracy, it was incorporated in the DfD system described here.

In this section the response of the filter Gp_2 has been modelled by considering the error between the discrete theoretical M/P ratio, $\frac{M(f_r; \alpha)}{P(f_r; \alpha)}$ and the Linear Model, $\frac{M'(f_r; \alpha)}{P'(f_r; \alpha)}$. As with $G^p m_1$, the proposed design uses the discrete M/P space, and a 1D prototype $G^p p_2$ was initially designed using the coefficient values of a cubic function. So,

$$Error(f_r; \alpha) = \frac{M'''(f_r; \alpha)}{P'''(f_r; \alpha)} = abs \left[\frac{M(f_r; \alpha)}{P(f_r; \alpha)} - \frac{M'(f_r; \alpha)}{P'(f_r; \alpha)} \right] = \frac{G^p p_2(f_r)}{G^p m_1(f_r)} \beta^3 \quad --(6)$$

It can be inferred that $\frac{G^p p_2(f_r)}{G^p m_1(f_r)} \beta^3$ can be modelled as a cubic function: $y = Cx^3$, where the gradient C at a particular discrete radial frequency f_r corresponds to the ratio $\frac{G^p p_2(f_r)}{G^p m_1(f_r)}$, hence by computing $C(f_r)$, and knowing the prototype $G^p m_1(f_r)$, the frequency response of a 1D prototype $G^p p_2$ can be determined. To compute the gradient, $y = Cx^3$ was fitted to the absolute error function and the prototype modelled as $G^p p_2(f_r) = C(f_r) G^p m_1(f_r)$. The 2D response and the filter coefficients, $gp_2(x, y)$ were then found in the same way as for the other filters.

2.2. Model Verification

The designed model was verified by working backwards to determine how well the designed filters fitted the theoretical M/P ratio. To enable an accuracy comparison with Watanabe's filters, the kernel size (7x7) and the number of frequency samples (32x32) were chosen as in [4]. As Watanabe [4] have not published a verification of their designs, and the numerical results for their 32 x 32 frequency responses were not available, direct comparison of the results was not feasible at this stage. Full comparisons can be made only on the estimated depth maps as described later. However a rough comparison with their filters was done by transforming their 7x7 spatial kernels into 32 x 32 frequency characteristics. These results have been presented in Section 3.

In the verification process, the frequency band up to which the M/P ratio remained monotonic was determined. In this section, the results have been based on the defocus condition:

$$\frac{e}{F_e} = 2.307 \text{ pixels}, \text{ that was used by Watanabe [4] when they obtained their filters. Using the}$$

equations in the Appendix, the following constants were calculated for the experiment:- $e = 17.746$ pixels, $F_e = 7.6923$, maximum blur diameter = 4.1614 pixels, focal length, $f = 50\text{mm}$, kernel size, $k_s = 7$ pixels, and aperture, $d = 6.5\text{mm}$. Again, by using the equations in the Appendix, the pattern frequency range used by the DfD calculation was $0.2857 \leq f_r \leq 0.3164 \text{ pixels}^{-1}$. A Matlab program was written to plot the theoretical M/P ratio, and the linear and error corrected models for a range of frequencies and normalised depth values, as shown in Figure (3). The mean square error estimate

between the theoretical M/P ratio and both the linear and the error corrected models have been provided in Table 1. It can be inferred that the filters devised by the new method fit well with the theoretical ones. More results for both simulated and real images have been presented in later sections.

Radial frequency (pixels ⁻¹)	MSE between M/P ratio and Linear Model	MSE between M/P ratio and Error Corrected Model
0.3141	0.0703	0.0636
0.3125	0.0630	0.0533
0.3078	0.0499	0.0397
0.2965	0.0296	0.0266

Table 1. Comparison of MSE between the Linear and Error Corrected Models

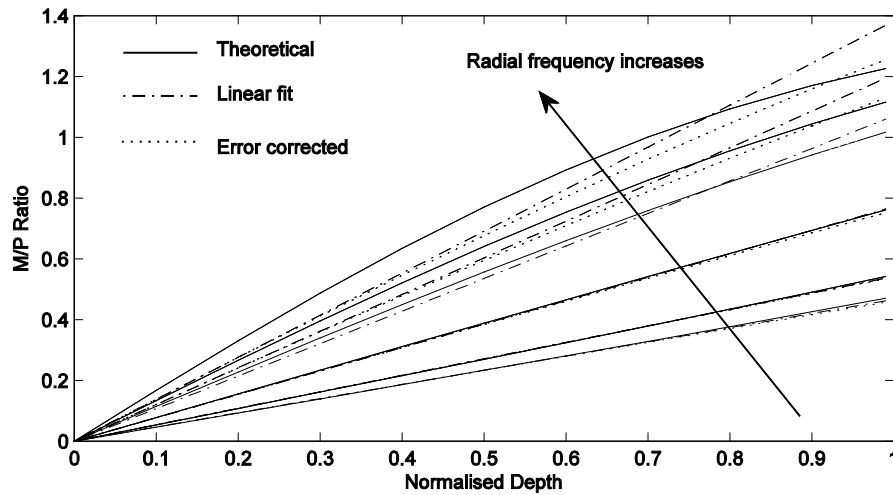


Figure 3. Model verification

3. Comparison with the Previous Filters

This section provides a comparison between the filters designed by the Two Step Polynomial approach presented in Section 2 with those designed by Watanabe [4]. The 7x7 filter coefficients of

both models were transformed into their equivalent 32x32 sample frequency responses, and verified as to how well they fit the theoretical M/P ratio. Figure (4a) shows the theoretical M/P ratio, Watanabe's Model [4], and the Two Step Polynomial results. The maximum texture frequency applicable for the defocus condition $\frac{e}{Fe} = 2.307 \text{ pixels}$ has been shown as it was the most curved.

Visually the proposed filters have provided a better fit. For a quantitative comparison figure (4b) shows the accumulated RMS errors for Watanabe's [4], and the Two Step Polynomial filters for all the frequencies within the applicable range. The RMS error was significantly lower for the Two Step Polynomial method particularly for normalised depths approaching one.

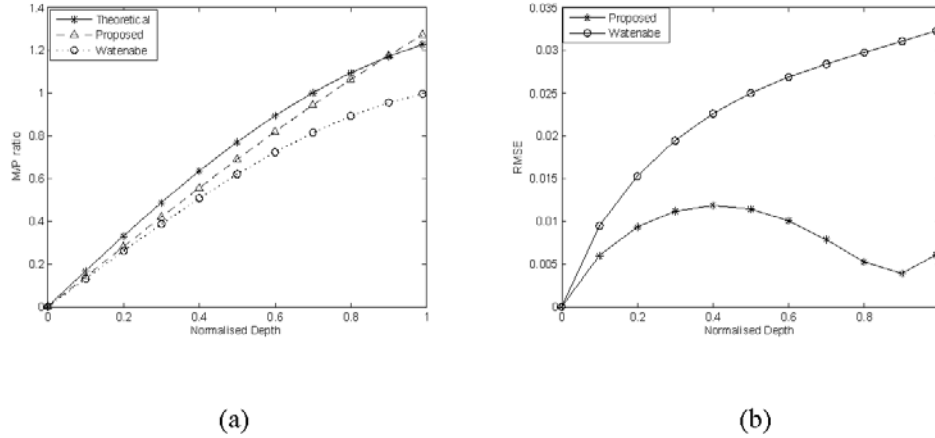


Figure 4. (a) Normalised depth vs. Theoretical M/P ratio for both models, (b) RMSE between theoretical M/P ratio and filter models

Next the normalised gain magnitude response of the filters was compared. Taking 1D slices through the 2D responses, as shown in Figure (5), filters Gm_1 and Gp_1 are similar for each model. However, there is a considerable dissimilarity in the response of the correction filter Gp_2 . It is noted that Gp_2 designed by the new model has a sharper transition between pass and stop bands, and a higher DC magnitude compared to Watanabe and Nayar's. This DC does not propagate in the depth estimation since the pre-filter suppresses any frequency response below the minimum cut-off frequency. Moreover the pre-filter designed by the new method has a smooth roll-off in the transition band compared to a sharp transition for Watanabe's. A sharp transition can propagate a ringing effect [23].

For a planar object that is perpendicular to the optical axis, the DfD should calculate a 2D depth map in which the depth estimates to each point on the plane are identical. This will be the case if the responses of the 2D filters are circularly symmetric. To compare the responses of the DfD systems constructed using both sets of filters a 2D image of an 8 bit sinusoidal intensity image of 400x400 pixels was simulated. The pattern was circular, centred on pixel (200,200), and of wavelength $\lambda = 3.2$ pixels. By smoothing the test image by convolution with the psfs of the lens in two stages of defocus, a pair of defocused images were obtained with which to test the DfD systems. The psfs were chosen to set a normalised depth $\alpha = 0.99$. The depth maps generated using both the models

have been shown in Figure (6). The mean depth error and standard deviation for the Two Step Polynomial model was 0.0454 and 0.0128, and for Watanabe's model 0.3615 and 0.2008 respectively. From Figure 6 and the standard deviation results it can be inferred that the depth map generated by the new filters is smoother and more planar than Watanabe's. This is because Watanabe's filters are not as circularly symmetric as the new set.

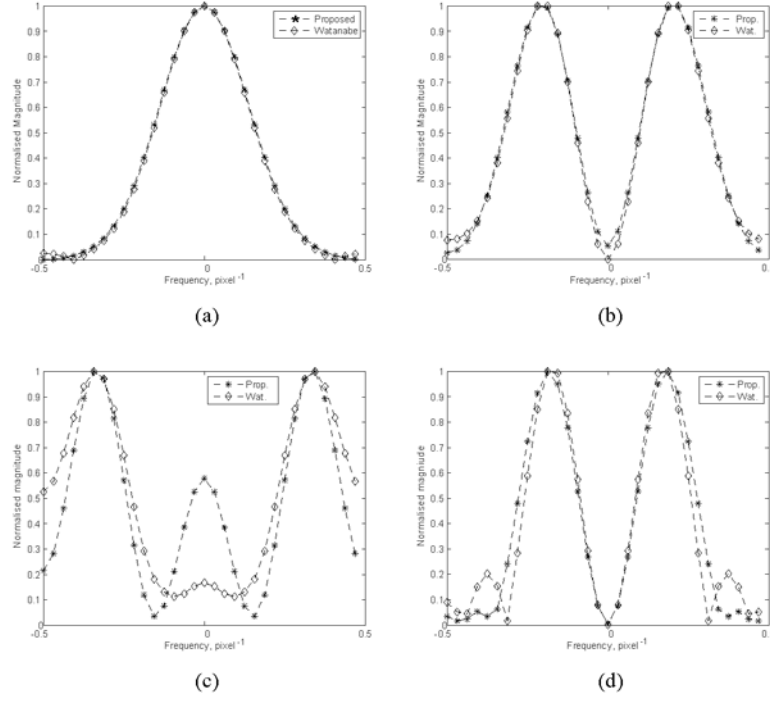


Figure 5. Magnitude responses of (a) G_{m1} , (b) G_{p1} , (c) G_{p2} and (d) Pre-filter

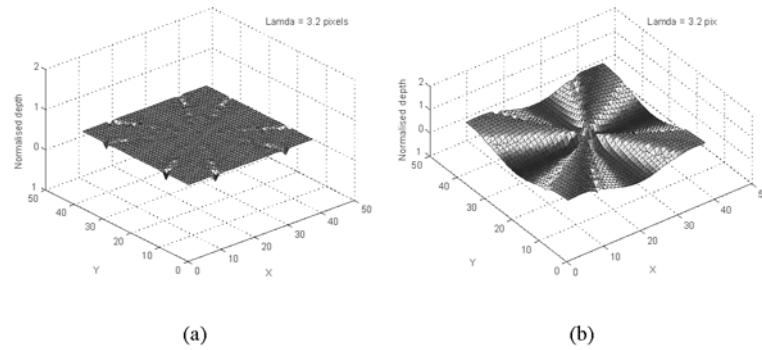


Figure 6. Depth map estimated using (a) The proposed filters, (b) Watanabe's filters

4. Depth Estimation Experiments

This section describes the results from a series of experiments designed to measure the depth accuracy of the DfD calculation when the new filters are used, and to compare them with Watanabe's published set [4]. The DfD algorithm described in [4] was coded and used to estimate the depth in each case. A full description of the algorithm can be found in [4]. For our particular implementation we smoothed the recovered depth map from each calculation using a 9x9 median filter as a post-process.

The experiments used both simulated and real images, each with 256 grey levels and a resolution of 400x400 pixels. They were selected to enable estimates of both the depth accuracy over a range of distances, and the dependence of the depth accuracy on the characteristics of the surface pattern of the object being measured. To enable a comparison the same input images were used in testing both the proposed filters and Watanabe's. Section 4.1 concerns simulated images. Firstly single spatial frequency patterns were used to render the surface of a 3D staircase object and the depth accuracy estimated for each step. Then a more complex and spatially varying pattern was used to render each step. In Sections 4.2 and 4.3 tests on naturally occurring textures and a range of actual objects from flat planes to composite objects have been reported.

4.1. Experimental Results with Simulated Images

In order to verify the DfD calculations for each of the two sets of filters used, sinusoidal brightness patterns with a single spatial frequency and different normalised depth values were developed. To enable a comparison with Watanabe [4] a defocus condition of 2.307 pixels was used, which limited the usable spatial frequency range to $0.2857 \leq fr \leq 0.3160 \text{ pixel}^{-1}$ and therefore the wavelength to $3.2 \leq \lambda \leq 3.5 \text{ pixels}$. The normalised depth range used was from 0.1 to 0.99. To simulate a depth staircase, the single frequency test pattern was defocused using a sequence of pillbox psfs in a way that every 40 pixel along the horizontal axis there was a step change in depth. This simulation enabled the estimated depth map to be viewed as a 3D staircase structure. Experiments were performed on two pairs of test images with pattern wavelengths of $\lambda = 3.2$ and 3.5 pixels . The linearity of the depth estimated by the two sets of filters has been compared in Figure (7a) and Figure (7b). It can be inferred that for wavelength $\lambda = 3.5 \text{ pixels}$, the depth map estimated by both sets are reasonably linear but for the lower wavelength (higher frequency) of 3.2 pixels the filter coefficients designed by the Two Step Polynomial approach provided a smoother and more accurate fit to the actual depth than Watanabe's. This increase in accuracy can be attributed to the new two step polynomial model which fits more closely to the theoretical M/P ratio as demonstrated in Figure (4). The statistics presented in Figure (7b) were calculated from local areas of $17 \times 371 \text{ pixels}$ which fitted well along each individual depth step. The standard deviation of the depth estimates was much lower when the new filters were used.

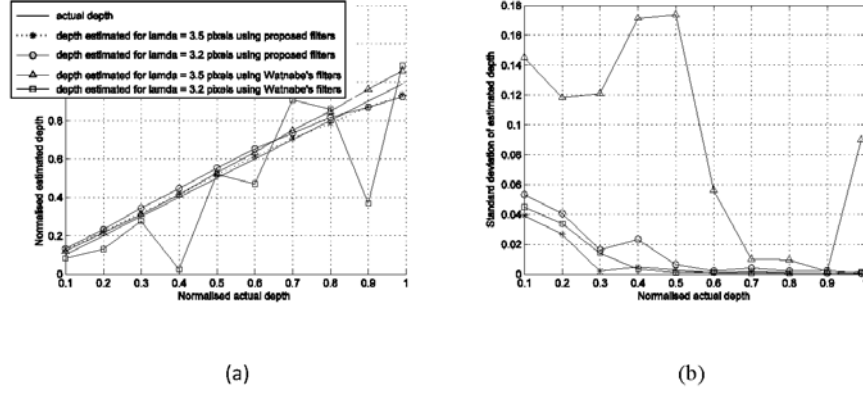


Figure 7. (a) Actual vs. Estimated depth at various normalised depths using the proposed and Watanabe's filters, (b) Standard deviation at these depths for both filter models

To verify the invariance of the filter coefficients to the image texture, a pattern set devised by Watanabe [4] that contained several differently texture stripes was used. Seven stripes had patterns with narrow spectral densities (PSD) centered on differing frequencies, and the remaining three had wide PSD patterns. The original pattern set was defocused using the pillbox psf to simulate a 3D staircase structure.

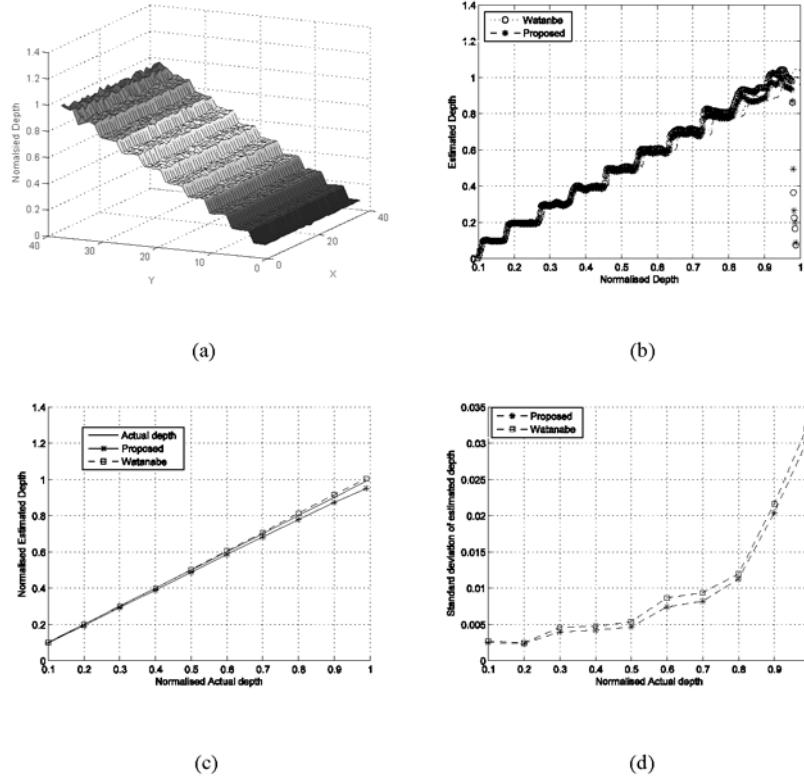


Figure 8. (a) 3D view of the estimated depth, (b) 1D plot of the estimated depth, (c) Actual vs. estimated depth for filters designed by both the models, (d) Standard deviation at different depths for both the models

Figure (8a) shows the estimated depth in 3D of the staircase using the new filters, and Figure (8b) shows a randomly section through the staircase for the filters designed by the new method and for those designed by Watanabe's. The linearity of the depth estimates for both the filter sets has been compared in Figures (8c) and (8d). It can be inferred that the filters designed by the new method are invariant to texture and provide a slightly better fit to the actual depth than Watanabe's. The relatively poor results from the higher steps are due to the use of the lower frequency section of the pattern there. In the following sections, experiments have been reported using real images and the accuracy of the depth estimated using both sets of filters has been compared.

4.2. Experiment with a random textured natural pattern: Abrasive Paper

This section provides depth estimation results using defocused images of a sheet of glass paper in a plane perpendicular to the optical axis. Eventually this pattern also served as the reference pattern used to calibrate the system since the PSD was wide within the working frequency range for the defocus condition. To enable a useful accuracy comparison with Watanabe [4], the defocus condition was again set to $\frac{e}{Fe} = 2.307 \text{ pixels}$. Using the equations in the Appendix, the working range was calculated to be 56mm, this was quite short but was limited by the pixel size of the camera and the aperture used. A larger pixel size or a narrower aperture would have provided an increase in the range [17].

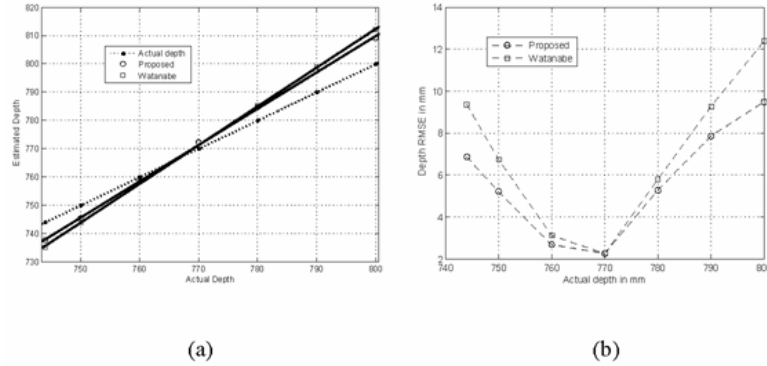


Figure 9. (a) Actual vs. Estimated distance (mm), (b) RMSE vs. Actual distance (mm)

To determine the accuracy of the setup, the glass paper pattern was moved along the optical path over the range 744mm to 800mm, with a pair of defocused images recorded and processed at every 10mm interval. The normalised mean depth was mapped to real world coordinates using the lens law. The depth estimation results for the two sets of filters were compared and are shown in Figure (9a). The RMS error plot at each distance has been presented in Figure (9b). From the plots it can be seen that the depth estimates using both the filter sets are reasonably linear. The RMS errors for the new filters were 6.8717mm and 9.489mm at the nearest and furthest distance respectively. This corresponds to an error of 0.9236% and 1.186% with respect to the centre of the lens to far-focus distance compared to 1.258% and 1.547% for Watanabe's filters. From these results it can be inferred that the filters designed using the Two Step Polynomial approach provided an improved

accuracy over Watanabe's for these natural textures. The next section provides depth results for actual 3D objects with natural textures.

4.3. Experiments with 3D objects and Natural textures

To complete the evaluation of the new set of filters, two real 3D objects with natural textures were measured. The objects were a multi-step staircase structures made from 3 pieces of mild steel on a background and a "T" structure made from natural wood. Figure (10) shows an original image of each structure, and Figure (11) the estimated depth maps. The shapes of the staircase and "T" can be identified easily in the depth maps. However, there were gaps between the steps on the staircase that were in deep shadow and this has resulted in large errors along the edges of each step. The objects and background were illuminated by a large area light source that has resulted in some specular reflections from their surfaces and hence less smooth depth maps than those obtained for the abrasive paper. As with human and stereo vision, that also require texture patterns for depth recovery, the shadows and specular reflections have resulted in some problems for DfD. Image noise introduced in the sensors or by the camera electronics has been identified as a source of errors for some image processes [24], but for DfD such noise is rejected by the algorithm as there is no correlation between the noise in each of the differently defocused images captured. To prove this noise was reduced in each input image by averaging several images for each defocus condition. The noise in the depth map was found to be independent of image noise.

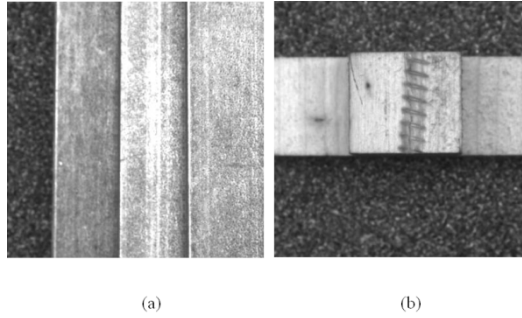


Figure 10. Near-focused images of (a) Staircase, and (b) "T"

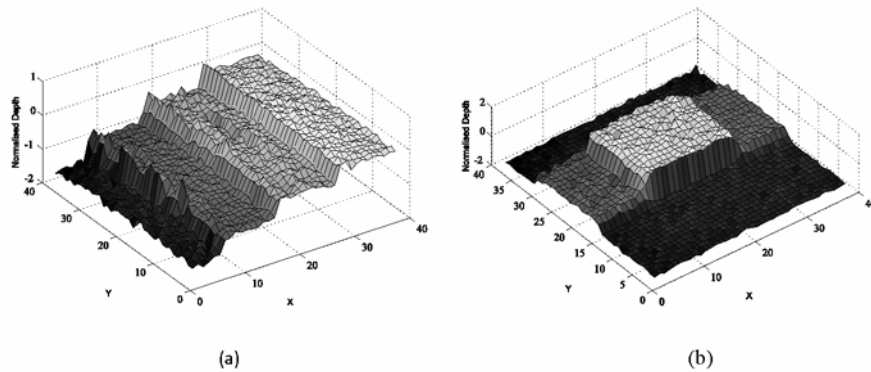


Figure 11. Estimated depth maps from (a) Staircase, and (b) "T"

5. Conclusion

The proposed design procedure based on the Two step polynomial model was simple to perform, and provided a better fit to the theoretical M/P ratios than Watanabe's filters [4] (Sections 2 and 3). Tests with simulated textures described in Section 3 proved that the two step polynomial filters were texture invariant and that they were more circularly symmetric than Watanabe's. The tests described in Section 4 showed that the filters designed by the new model provided a better fit to the actual depth than those of the previous design. Generally, tests with real images of arbitrary natural textures (glass paper) resulted in slightly higher errors than for simulated images (Section 4.2) as the actual lens psf was defocusing the images as opposed to the theoretical cylindrical shaped psf that was used to model the M/P ratios. For the filters designed by the Two Step Polynomial approach the RMS error with respect to the distance was 0.9236% at the near-focused plane and 1.186% at far-focused, compared to 1.258% and 1.547% for Watanabe's filters. From these results it can be inferred that the new filters estimated the depth at a higher accuracy than the previous filters. Moreover, the design procedure explained in Section 2 can be effectively applied for any defocus conditions by simply modelling the psf.

Appendix: Setting up and verifying the Working Distance for the Rational Filters designed by the Two Step Polynomial Approach

Referring to Figure 1, given the defocus condition, D , and the far-focused object distance u_f , the working distance is $u_f - u_n$, where u_n is the near-focused image distance. The f -number of the lens,

$F_e = \frac{f}{d}$. Where f is the focal length of the lens and d is the diameter of the aperture. The distance

between the near and far-focused images, $2e = 2 * D * F_e * pixsize$, where $pixsize$ refers to the camera CCD element size which was $7.4\mu m$ for our camera and $13\mu m$ for Watanabe's. The far-

focused image distance, $v_f = \frac{u_f * f}{u_f - f}$, and the near-focused distance $v_n = v_f + 2e$. Now $u_n = \frac{v_n * f}{v_n - f}$

.

In practice the working distance will be known and then the defocus condition can be calculated and

verified. Determine $e = \frac{v_n - v_f}{2 * pixsize}$ pixels and $D = \frac{e}{F_e}$ pixels. Now D must satisfy the following

constraints:- (a) The maximum frequency that can be resolved, $\max f_r = 0.73 * \frac{F_e * pixsize}{e}$ pixel⁻¹;

(b) The minimum frequency that can be resolved by the filter, $\min f_r = \frac{2}{k_s}$ pixel⁻¹, where k_s is the

size of the kernel (7 pixels); The factor of 0.73 comes from the maximum blur circle diameter, $\frac{2e}{F_e} \leq 0.73k_s$ pixels and ensures a monotonic M/P ratio.

References

- [1] P. Grossmann, Depth from focus, *Pattern Recognition Letters*, 5 (1) (1987) 63-69.
- [2] Aamir Saeed Malik and Tae-Sun Choi, A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise, *Pattern Recognition*, 41 (7) (2008) 2200-2225.
- [3] A.P. Pentland, A new sense for depth of field, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9 (4) (1987) 523-531.
- [4] M. Watanabe and S.K. Nayar, Rational filters for passive depth from defocus, *International Journal of Computer Vision*, 27 (3) (1998) 203-225.
- [5] M. Subbarao and T.C Wei, Depth from defocus and rapid auto focussing: a practical approach, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, USA, 773-776, 1992.
- [6] M. Subbarao and G. Surya, Depth from defocus: spatial domain approach, *International Journal of Computer Vision*, 13 (3) (1994) 271-294.
- [7] J. Ens and P. Lawrence, An investigation of methods for determining depth from defocus, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15 (12) (1993) 97-108.
- [8] O. Ghita and P.F. Whelan, Real time 3D estimation using depth from defocus, *Proc. of Irish Machine Vision and Image Processing Conference*, Dublin City University, Ireland, 167-181, 1999.
- [9] A.N. Rajagopalan and S. Chaudhuri, An MRF model-based approach to simultaneous recovery of depth and restoration from defocused images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21 (7) (1999) 577-589.
- [10] A.N. Rajagopalan and S. Chaudhuri, *Depth from defocus: a real aperture imaging approach*, Springer, New York, USA, 1999.
- [11] P. Favaro and S. Soatto, A geometric approach to shape from defocus, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 (3) (2005) 406-417.
- [12] P. Favaro and S. Soatto, Learning shape from defocus, *Proc. of 7th European Conference of Computer Vision*, Copenhagen, Denmark, Part 2, 735-745, 2002.
- [13] F. Deschenes, D. Ziou, P. Fuchs, Improved estimation of defocus blur and spatial shifts in spatial domain: a homotopy-based approach, *Pattern Recognition*, 36 (9) (2003) 2105-2125.
- [14] F. Deschenes, D. Ziou, P. Fuchs, A homotopy-based approach for computing defocus blur and affine transform simultaneously, *Pattern Recognition*, 41 (7) (2008) 2263-2282.
- [15] D. M. Tsai and C. T. Lin, A moment preserving approach for depth from defocus, *Pattern Recognition*, 31 (5) (1998) 551-560.
- [16] Li Ma and R. C. Staunton, Integration of multiresolution image segmentation and neural networks for object depth recovery, *Pattern Recognition*, 38 (7) (2005) 985-996.

- [17] A.N. Joseph Raj, Accurate Depth from defocus estimation with video rate implementation, Ph.D. Dissertation, Univ. of Warwick, Coventry, U.K. Dec 2009.
- [18] M. Watanabe and S.K. Nayar, Telecentric optics for focus analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19 (12) (1997) 1360-1365.
- [19] A.N.J. Raj and R.C. Staunton, Estimation of image magnification using phase correlation, Proc. International Conference on Computational Intelligence and Multimedia Application, Sivakasi, India Vol. 3, 490-494, 2007.
- [20] T.J. Schultz, Multiframe image restoration, in *Handbook of Image and Video Processing*, Al Bovik ed. Academic Press, San Diego, CA, USA, 175-189, 2000.
- [21] C. D. Claxton and R. C. Staunton, Measurement of the point-spread function of a noisy imaging system, Journal of the Optical Society of America A, 25 (2008) 159-170.
- [22] D. Marr and E. Hildreth, Theory of edge detection, Proc Royal Society London, B207 (1980) 187-217.
- [23] Jae S. Lim, Two Dimensional Signal and Image Processing, Prentice Hall Inc, New Jersey, USA, 1990.
- [24] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 2nd Edition, Prentice Hall Inc, New Jersey, USA, 2002.